



**QUEEN'S  
UNIVERSITY  
BELFAST**

## 临床技能竞赛试题分析与计算机实现

Liu, J., Zhu, Y., Qu, B., Chen, T., & Yao, J. (2016). 临床技能竞赛试题分析与计算机实现. *China Higher Medical Education*, 89-90. <https://doi.org/10.3969/j.issn.1002-1701.2016.04.046>

**Published in:**  
China Higher Medical Education

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

**Publisher rights**  
© 2015 China Higher Medical Journal

**General rights**  
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**  
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

# 临床技能竞赛试题分析与计算机实现

刘洁<sup>1</sup>, 朱亚鑫<sup>2</sup>, 曲波<sup>2,\*</sup>, 陈天豹<sup>3</sup>, 姚江<sup>3</sup>

<sup>1</sup> 中国医科大学公共卫生学院

<sup>2</sup> 中国医科大学医学教育研究中心

<sup>3</sup> 贝尔法斯特女王大学药学院

\* 通讯作者

[摘要] 在医学考试过程中, 试题的质量直接影响着考试的整体水平, 做好试题评价和分析对提高试题质量, 优化更新试题库尤为重要。文章以“全国高等医学院校大学生临床技能竞赛——东北、华北赛区”学生比赛的试题为例, 详细阐述了试题的难度、区分度和信度等指标的意义, 并通过计算机统计软件实现, 为高等医学院校的教师进行试题评价提供了参考。

[关键词] 难度; 区分度; 信度; 统计软件; 试题评价

医学考试的主要目的是对考生的医学素质和能力作出评价, 因此试题的优劣直接影响着整个考试的质量。试题评价不仅为教师的教学及命题提供有价值的信息, 并且还还为优化和更新试题库, 进行科学管理提供参考<sup>[1]</sup>。本文以“全国高等医学院校大学生临床技能竞赛——东北、华北赛区”学生比赛的试题为例, 介绍了试题评价的各种指标及其计算机实现方法, 为教学研究和管理人员进行试题评价及如何提高命题质量提供参考。

## 一、 试题评价常用指标介绍

### 1. 难度。

难度(Difficulty) 是表示试题难易程度的指标。对于主观题来说, 难度为所有考生在该题的平均得分与该题的满分之比。对于客观题来说, 一般其结果只有正确与错误之分, 这时, 试题的难度用答题的通过率来表示。难度值越大, 表示此题越容易。

### 2. 区分度。

区分度(Discrimination) 是指试题或试卷对考生实际水平的区分程度, 常用 D 来表示<sup>[2]</sup>。从概念上来说, 试题的区分度可以被理解为试题难度与考生能力间的关系。它基于一个假设, 即: 高能考生比低能考生更有可能正确回答试题。一般用所有考生的某题得分与所有考生的考试总分之间的相关系数 r 来表示考试中某一题的区分度, 试题的区分度在 0.4 以上表明此题的区分度很好<sup>[3]</sup>。

### 3. 信度。

信度(Reliability) 是反映考试稳定性和可靠性的指标<sup>[2]</sup>。信度有很多种, 目前比较常用的是内部一致性信度(Internal consistency reliability), 用来测验试卷内部所有试题的一致性。常用克朗巴赫(Cronbach) 系数  $\alpha$  来表示, 美国医学考试委员会曾规定: 只有信度  $> 0.70$ , 该考试所得分数才可以用于评价个人和集体; 如果是用来鉴定应试者个人的考

试，信度必须达到 0.90 或更高<sup>[4]</sup>。

上述几个指标中，难度表示试卷的难易程度，区分度可以用来说明考生能力与试题难度的关系，信度反映考试的稳定程度，是保障试题质量的基础。他们既有各自独立的意义，又有相互联系制约的一面。不能简单认为，当试题难度大时，就能很好地区分考生的能力。在实际工作中必须根据考试目的和教学实际等情况，综合考虑以上三个指标的得分情况，才可以为试题评价者提供更科学的信息<sup>[5]</sup>。

二、实例分析

本研究将“全国高等医学院校大学生临床技能竞赛——东北、华北赛区”的 28 支参赛队伍作为研究对象，28 支参赛队伍的比赛试题共 20 道，均为临床实践操作，比赛的满分为 100 分，每道题的满分情况具体如下（见附表）。研究以此次竞赛的试题为例，进行了试题评价，并详细介绍了 SPSS 19.0 的实现过程。

附表 临床技能竞赛 28 支队伍参赛的试题情况及分析结果（n = 20）

题号	试题满分	难度	区分度
Q01	2	0.491	0.772
Q02	4	0.438	0.871
Q03	4	0.375	0.879
...	...	...	...
Q19	2	0.482	0.620
Q20	8	0.545	0.693

1. 数据库的建立。

进入“数据视图”界面，将所需要的数据进行输入，最终数据库的建立结果如下（见图 1）。

队伍编号 1- 28 分别代表 28 支参赛队伍，变量“评分值”为每支队伍的参赛总得分，变量“Q01-Q20”为 1- 20 题的得分。图 1 仅展示部分数据。

2. 难度的计算。

打开 SPSS 数据库，选择“分析”→“描述性统计”→“描述性分析”命令，打开“描述性分析”主对话框（见图 2）。

队伍编号	评分值	Q01	Q02	Q03	Q04
1	49.5	1.5	3.0	.0	2.0
2	43.0	.5	.0	2.0	3.0
3	37.0	.0	1.0	.0	.0
4	68.5	2.0	3.0	3.0	3.0
5	49.0	1.0	.0	.0	2.0

图 1 数据库的建立



图 2 “描述性分析” 主对话框

将所要分析的变量，即“Q01-Q20” 移动到右侧的“变量(V)” 中，并点开左上角的“选项”，将其中的“均值” 选中，点击“确定”，结果中的最右侧一列表示每道题得分的均值，例如 Q01 题的均值为 0.982，由于 Q01 题的满分为 2 分，故其难度  $P1 = 0.982 / 2 = 0.491$ 。以此类推，剩余题的难度计算方法与 Q01 题的难度算法相同，各题的难度得分见附表。

### 3. 区分度的计算。

打开 SPSS 数据库，选择“分析” → “相关性” → “双变量相关性” 命令，打开“双变量相关性分析” 主对话框( 见图 3) 。



图 3 “双变量相关性分析”主对话框

将要分析的题“Q01”和“评分值”放入“变量(V)”中，点击“确定”，Q01 题和总分的 Pearson 相关系数为 0.772，因此 Q01 题的区分度  $D1 = 0.772 > 0.4$ ，结果表明第 1 题对考生水平的区分效果很好。其他试题区分度的算法与 Q01 题相同，最终结果见附表。由附表可以看出，本次竞赛中每道题的区分度值均大于 0.4，表明区分效果都很好。

#### 4. 信度的计算。

打开 SPSS 数据库，选择“分析”→“度量”→“可靠性分析”命令，打开“信度分析”主对话框（见图 4）。



图 4 “信度分析” 主对话框

将要分析的变量，即每个题目的得分“Q01-Q20”放入图 4“项目(I)”选项中，在“模型”选项中选择“ $\alpha$ ”，点击“确定”，克朗巴赫系数  $\alpha = 0.946$ ，此次临床技能竞赛试题的信度得分为  $0.946 > 0.9$ ，表明此次竞赛试题的信度较好。

综上所述，此次竞赛每道题的难度值介于 0.4-0.7 之间，区分度得分均大于 0.4，试题整体的信度值为 0.946。因此，从各项评价指标来看，该竞赛试题的质量水平较好。

### 三、 小 结

经试题分析，我们发现此次竞赛试题信度、难度和区分度非常好。教育测量在检验教学效果、评估学业成绩等方面的应用极为广泛，熟悉并掌握基本的试题评价技术是高等学校教师应该具备的基本技能<sup>[1]</sup>。关于试题评价的指标还有很多，本文只是对几个主要的指标进行了介绍。近年来，随着计算机技术的不断发展和应用，已经将教育工作者从手动统计和分析数据的繁琐工作中解放出来，使试题评价工作更加快捷、准确，在医学领域有着广阔的应用前景。

### 【参考文献】

- [ 1 ] 王 渊 . 考试质量分析系统的设计 [ J ] . 医学教育探索 , 2010 , 9( 7 ) : 971- 974.
- [ 2 ] 田考聪 , 彭 斌 . 试卷质量定量分析系统中的几个参数及其应用 [ J ] . 医学教育探索 , 2004 , 3( 4 ) : 52- 54.
- [ 3 ] 陶 英 , 林佩艳 . Excel 数组公式在区分度计算中的应用 [ J ] . 中国教育技术装备 , 2009( 18 ) : 134- 135.
- [ 4 ] 刘叔才 , 葛利荣 . 医学考试质量评价指标体系的构建与实现 [ J ] . 中国社会医学杂志 , 2008 , 25( 4 ) : 202- 204.
- [ 5 ] 吴珍珠 . 探讨试题难度与区分度的关系 [ J ] . 教育教学论坛 , 2011( 30 ) : 235- 237.

**【作者简介】** 刘洁,女,博士,副教授,研究方向:医学教育测量与评价;卫生统计方法学。

**【基金项目】** 2014 年辽宁省普通高等教育本科教学改革研究项目拟立项项目(编号: UPR P20140243, UPR P20140976)。